DOCUMENT RESUME

ED 440 127 TM 030 739

AUTHOR Ediger, Marlow
TITLE Assessment with Portfolio and Rubric Use.
PUB DATE 2000-03-06
NOTE 9p.
PUB TYPE Opinion Papers (120)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Elementary Secondary Education; Performance Based Assessment; *Portfolio Assessment; *Scoring Rubrics; *Student Evaluation; Test Use

ABSTRACT

The use of portfolios and rubrics in assessment developed in part as a reaction to limitations of standardized tests for student evaluation. Portfolios relate directly to what students are learning, but there are many problems in evaluating their contents. These center on: (1) time involved in scoring; (2) lack of validity statistics; (3) relative absence of reliability data in comparison with standardized tests; (4) vagueness in reporting portfolio results; and (5) the necessity of different criteria for different products. To remedy these perceived weaknesses of the portfolio process, rubrics came into being to quantify student achievement. Rubrics, in the form of agreed-upon criteria, guide raters in assessment. Rubrics may be difficult to write and time consuming to use, but they make the portfolio evaluation process more accountable and support objective judgments. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 440 127

Assessment with Portfolio and Rubric Use

Marlow Ediger

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Marlow Ediger

BEST COPY AVAILABLE

TM030739

ASSESSMENT WITH PORTFOLIO AND RUBRIC USE

Portfolio and rubric use in assessment was a reaction to standardized and criterion referenced tests (CRTs) to ascertain and document student achievement in different curriculum areas. Advocates of portfolios and rubrics believe that testing and measuring procedures have the following weaknesses:

1. a single percentile is to "tell the story" about a student's achievement.
2. these tests, be they standardized or CRTs, are given once a year at the most.
3. they tell little abut specific kinds of errors students made individually and what con be done to remedy the deficiencies.
4. feedback is not available to parents to understand how to assist the involved offspring in the home setting.
5. each test item is isolated from the others. Students cannot see the relationship of subject matter in these situations.
6. multiple choice test items are used and these do not reflect doing practical things in everyday life's situations. Rather responding to paper/pencil test items is in evidence.
7. test items are written by those removed from the local classroom where teaching and learning occur.
8. test writers then cannot provide help to students as needed in ongoing lessons and units of study in the classroom.
9. contextual assessment is lacking since there is no attempt to assess within a specific activity where error(s) were made.
10. continuous assessment is not being stressed with a yearly assessment being in evidence.

Portfolios then attempted to care of selected vacuums that existed in standardized tests and CRTs (See Ediger, 1997, Chapter 16).

Portfolios and the Curriculum

Portfolios and their contents pertain to what students have had opportunities to learn. These learning opportunities stress what is written inside the stated objectives in the curriculum. Each student with teacher guidance is actively involved in choosing products and processes to go into a portfolio for the former. The following are examples of what might become a part of a portfolio:

1. written work from different academic areas.
2. a video tape showing the quality of interaction in committee work.
3. a cassette to indicate the quality of oral work done, such as in

reading and speaking endeavors.

4. snapshots of construction experiences as they relate to ongoing lessons and units of study.

5. drawings, diagrams, charts, and graphs developed by the learner.

6. art work in terms of collages, friezes, murals, and mobiles developed within a learning activity.

7. journal and diary entries to reveal what has been accomplished.

8. self evaluation forms that describe the learner's feelings about his/her achievement.

9. diverse dramatizations participated in to clarify concepts/generalizations, described in narrative or expository content.

10. summary statements on conferences held with the classroom teacher to assess progress (Ediger, 1997, 20- 26).

The size of the portfolio should not be too voluminous, nor too minimal. It should indicate what a student has learned to achieve objectives in the curriculum. Two to three professionals should be involved in assessing each portfolio. Inservice education is necessary for professionals in order for each to be a good evaluator of portfolios. There needs to be agreement among the evaluators for each portfolio so that interrater reliability is in evidence. Otherwise there may be little merit in having any one portfolio appraised if there is widespread disagreement as to its quality. If interrater reliability is low, then the questions arises as to how good a portfolio actually is. If two to three raters agree upon what rating to give a portfolio, then the reliability should be high in assessing if there is considerable agreement among the numerous portfolios.

Any assessment procedure tends to have its pros and cons. Portfolios are no exception. The criticisms for portfolio use come largely from those who advocate a measurement philosophy to ascertain student achievement. The first criticism pertains to the time and money needed to score portfolios. Teachers, no doubt, are asked to score portfolios. The chances are the time allotted here for evaluation will be great indeed! If there are twenty-five students in a classroom and each completed a portfolio, there may be 250 pages to assess, 25 X 10 pages each. this would be a nominal figure. If each student had twenty pages of portfolio content, then 500 pages would need to be assessed. If two teachers are to assess each portfolio in a class, then the previous number of pages would need to be doubled per teacher assessor. Both classrooms of portfolios need to be assessed. Something may have to give here in terms of time taken from teaching students. Or, midnight oil would need to be "burned" which may rob teachers of time for sleep and relaxation. Supposing outside assessors are to be hired, the cost could

then be very high indeed with the many hours necessary to assess portfolios in a school system. Standardized tests and CRTs are machine scored and mass numbers of student test results can be shown in a printout almost instantaneously (Ediger, 1994, 169-174).

Second, portfolios will not show validity statistics as compared to CRTs, in particular. When CRTS are used in a school, the accompanying objectives, if very precisely written, do provide the teacher with clear benchmarks for teaching. Thus, the test items on the CRT might well be clearly aligned with the stated objectives. If the teacher used the objectives, conscientiously, in teaching, validity of the test should indeed be quite high. With portfolios, it is more difficult to align its contents with more broadly written objectives. Within any portfolio, there is much to assess and attempt to pinpoint inherent products and processes with the broadly stated objectives. This is a monumental task.

Third, reliability data can be determined readily with standardized tests and CRTs. Test/retest, alternative forms. and split hall reliability figures have been worked out by companies producing these tests, especially in pilot studies. Students' test results from the local school/classroom may be compared with the norm group as given in the manual of the administered test. The printout of student test results will provide the necessary information in comparing the test taker's test results with the norm group. Standardized tests will have more data on validity and reliability as compared to CRTs developed on the state level. It costs much money to do pilot studies and publish the results. Commercial companies which publish standardized tests have more money available to run these studies as do statewide developed CRTS. Thus, with numerical results from testing, validity and reliability data may be computed quickly. Portfolios and their inherent contents prevent these specific comparisons on validity and reliability to be made.

Fourth, there is vagueness in portfolio results from assessors if the ratings of "poor," "average," "good," and "proficient" are given to students. The dividing lines are not clear and may make for arbitrary decisions in many cases.

Fifth, different products need diverse criteria for assessment. For example, the criteria used to assess the quality of a business letter are different than for a haiku poem. To be sure if the following are assessed in a written product, the results might always be consistent: spelling, punctuation in most cases, and proper capitalization of letters. However, creativity, such as in a haiku poetry, is very difficult to ascertain. What is creative behavior for one student may be rote learning for another (See Salvia and Ysseldyke, 1995).

Using Rubrics to Assess

To remedy the five selected weaknesses listed above, rubrics

came into being to quantify achievement of students, particularly from assessing the portfolio. Rubric advocates attempt to make the assessment of portfolios as objective as possible. How is this to be done? Ratings may be given any product in a portfolio from one to seven. Fewer categories in the rating scale may also be used such as from one to three, depending upon what is meaningful to raters in differentiating student's products and processes. Within each level, there are comments, carefully developed and written, to indicate what a student needs to be able to do to achieve level one, two, or three. The following are criteria to use in assessing student achievement in writing a business letter:

Level one---- contents written lack clarity, details are lacking, too many spelling errors, slovenly work inherent, improper form used for the diverse parts of the business letter.

Level two---- contents written are clear, details need expansion, contains a few spelling errors, shows considerable effort in writing, most of the parts of the business letter are in proper form.

Level three---- contents written possess clarity, necessary specific information is contained in the letter, no spelling errors, superb job of writing, all parts of the business letter are intact in the written product (See Skillings and Ferrell, 2000).

The above agreed upon criteria provide guidance to the rater as to how to assess a student's portfolio. Criteria listed might even be more specific such as giving weight to each standard within a specific level, such as how much emphasis should correct spelling of words receive as compared to having the different parts of the business letter properly spaced and included.

There are numerous advantages given for the use of rubrics to assess portfolios more objectively than would otherwise be the case.

1. numerals are provided after assessing a given product/process. By assessing in terms of criteria such as for levels one to three, raters have a better chance of agreeing upon the student's level of achievement. The criteria provide the needed guidance in providing the rating.

2. students, after observing the different criteria on a rubric have a better idea of what is expected in order to receive top ratings. Higher achievement may then be in the offing. There are schools where students have assisted in writing criteria for the different levels of achievement in a rubric.

3. content in the rubric relates directly to what should be contained in quality products/processes produced by learners.

4. assessment is then tied into the instructional procedures within a classroom. Outside "experts" such as test corporations are then not involved in assessing student achievement.

5. students may then readily receive rubric feedback that involves opportunities to improve over previous attempts from portfolio results.

6. learners may be provided opportunities to self assess their work against rubric standards.

7. raters' results from rubric use may be compared to ascertain reliability.

8. rubric criteria may be changed, as needed, to increase validity to cover what has been taught. Face validity is used in the process.

9. rubrics may standardize the criteria used to assess student work.

10. rubrics are an attempt to standardize portfolio results with the use of numerals or levels to indicate where a learner is presently in achievement (See Walsh and Betz, 1985).

As is true of most ideas, plans, and procedures, there is an opposite and equal reaction to rubric use. These reactions are the following:

1. the time given to develop and use rubrics in scoring portfolios.

2. the subjectivity in scoring portfolios, even when using rubrics. Rubric criteria for each level are subject to interpretation. Perhaps, this is no less subjective here than test writers determining which subject matter items students should be tested upon in a standardized test or CRT.

3. numerals can become too important to the hearer of the portfolio results, as compared to the actual specifics contained in the student's product/process.

4. many, many rubrics may need to be written since each kind of product/process may differ from another such as assessing quatrain poetry as compared to expository writing in a portfolio.

5. considerable inservice time needs to be given to the proficient use of portfolios and rubrics. However, preservice and inservice education is necessary in so many facets of teaching and learning. How many years would it take to educate in all necessary areas of teacher education in schools of education? There are many, many recommendations in journal articles and speeches given at teacher education conventions in terms of what to do to improve university experiences for students. Sometimes, there is considerable disagreement as to what to do to improve the public schools. For example, behaviorism with its measurement philosophy and humanism with its student input into curriculum development are far apart when thinking of what needs to be done to improve student's educational experiences in the public schools.

Conclusion

Assessing student achievement and reporting results to responsible others is indeed a complex process. There are numerous issues that need resolving in the assessment controversies. The following appear to be relevant issues in assessment:

1. How objective can an assessment instrument be to ascertain learner progress? This brings one's attention the standardized test/CRT versus portfolio disagreements.
2. How much of student achievement can be measured to obtain precise numerical results such as percentiles, as compared to more openended data such as in portfolios and rubrics?
3. How much information is needed from each student to learn about the individual's progress? Here, the issue is a single test score as compared to daily student results from ongoing lessons and units of study.
4. How frequently should students be assessed? The issue here pertains to annual assessments in standardized testing/CRTS as compared to rather continuous evaluation such as in daily teacher/student appraisals with portfolio/rubric documentation.
5. Who should be involved in determining how students should be assessed?
6. How should accountability of teachers be determined? The inherent issue involves accountability of classroom teachers as well as writers and publishers of standardized tests/CRTs.
7. Should standardized tests/CRTs be open to scrutiny to indicate subject matter on these tests for responsible individuals to view and assess?
8. How accountable are publishers of tests in which the profit motive can be very strong? Lobbying is a powerful factor in determining prices and needs in a market economy. Hardly does the law of supply and demand enter in to determining prices of commodities.
9. Is it possible to truly document student learning when the conditions of testing and evaluation enter in to arriving at the final test score, such as a percentile? Here, the writer is reminded of taking a standardized test on the graduate level where the room was hot/crowded indeed and a test taker sat nearby continually chewed gum loudly in an obnoxious way.
10. How can measurement philosophies be harmonized with contextualism to determine student achievement?

References

Ediger, Marlow (1997), "Appraising Student Achievement in Mathematics," Teaching Mathematics in the Elementary School, Kirksville, Missouri: Simpson Publishing Company, Chapter 16.

Ediger, Marlow (1997), "Portfolios, Pupils, and Their Teacher," Education Magazine, nr. 120, published by the Qatar National Commission for Education, Doha, Qatar (the Middle East).

Ediger, Marlow (1994), "Measurement and Evaluation," Studies in Educational Evaluation, 20 (2), 169-174.

Salvia, John, and James F. Ysseldyke (1995), "Using Student Portfolios in Assessment," Assessment, Sixth Edition. Boston: Houghton Mifflin Company, Chapter Twelve.

Skillings, Mary Jo, and Robbin Ferrell (2000), Student Generated Rubrics: Bringing Students into the Assessment Process," The Reading Teacher, 53 (6), 452- 456.

Walsh, W. Bruce, and Nancy E. Betz (1985), Tests and Assessment. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.